

TRIBHUVAN UNIVERSITY

Institute of Science and Technology

2066

Bachelor Level/ First Year/ First Semester/ Science
Computer Science and Information Technology (Stat. 108)
(Statistics I)

Full Marks: 60
Pass Marks: 24
Time: 3 hours.

Candidates are required to give their answers in their own words as far as practicable.
All notations have the usual meanings.

Group A

Attempt any two:

(2x10=20)

1. Describe simple random sampling with and without replacement for drawing a random sample of size n from a population of size N . In both cases show that sample mean is unbiased estimate of the population mean. Derive the variance of the sample mean in both cases. Show that the ratio of the variance of sample mean in sampling without replacement to that sampling with replacement is $\frac{N-n}{N-1}$. Comment on this result.
2. Describe the function and the procedure of the Kruskal-Wallis one-way ANOVA test by ranks. The original data of three independent samples were collectively converted to ranks as shown in the adjacent table. Set up appropriate null and alternative hypotheses and carry out the Kruskal-Wallis test at 5% level.

Sample 1	Sample 2	Sample 3
1	2	5
4	3	9
6	7	12
10	8	18
11	13	20
14	15	22
16	17	23
19	21	24

3. In order to establish the functional relationship between annual salaries (Y), years of education past high school (X_1), and years of experience with the firm (X_2), data on these three variables were collected from a random sample of 10 persons working in a large firm. Analysis of data produces the following results
 - (a) The total sum of squares, $\sum(Y_i - \bar{Y})^2$, is 397.6 and sum of squares due to error $\sum(Y_i - \hat{Y}_i)^2$, is 23.5. Compute the value of R^2 and interpret the result. Also compute the value of F statistic for testing the significance of the model.
 - (b) The estimated regression model of Y on X_1 and X_2 is presented below where the figures within parentheses are the standard errors of the estimated regression coefficients. Interpret the model and carry out the test of significance of the two

slope regression coefficients at 5% level by stating the null and the alternative hypotheses explicitly.

$$\hat{Y} = -8.883 + 1.85X_1 + 2.92X_2$$

(4.94) (0.59) (0.61)

Group B

Answer any eight questions:

(8x5=40)

4. Show that sample variance in simple random sampling method is an unbiased estimator of population variance.
5. Write the sample multiple linear regression model of Y on X_1 , X_2 and X_3 based on a sample of size n. What are the assumptions to be made on this model for estimation and test of significance?
6. Test whether the color of sons' eyes is associated with that of the fathers at 5% level of significance using the data available in the following table.

	Sons' eye color		Row Total
	No light	Light	
Father's eye color			
Not light	230	148	378
Light	151	471	622
Column total	381	619	1000

7. Define the first order autocorrelation? Estimate the first order autocorrelation from the data available in the table below.

Time (t)	1	2	3	4	5	6	7	8	9	10	11
Residuals ($\hat{u}_t$)	-5	-4	-3	-2	-1	0	1	2	3	4	5

8. In the stratified sampling sing simple random sampling without replacement method in each stratum, show that $Var(\bar{y}_n) = \sum \left(\frac{N_h}{N} \right)^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_h^2}{n_h}$. Simplify this expression when the total sample of size n is allocated according to proportional allocation across strata.
9. To evaluate a speed reading course, a group of 10 subjects was asked to read two comparable articles – one before the course and one after the course. Their scores on reading test are follows:

Before course (x)	57	80	64	72	90	59	76	98	70	83
After course (y)	60	90	62	79	95	58	80	99	75	94

Test whether the course is beneficial using the Wilcoxon Signed Rank test at 5% level of significance, given that $P(T^+ > 44|H_0) = 0.042$ where T^+ is the sum of the positive ranks of the difference $d_i = (y_i - x_i)$.

10. Define multiple regression coefficient between a variable X_1 and a set of variables $X_2, X_3, \dots, X_4$. If $r_{12} = 0.85$, $r_{13} = 0.75$, and $r_{23} = 0.55$, then compute the multiple correlation coefficient between X_1 and X_2, X_3 , where r_{ij} is the simple correlation coefficient between X_i and X_j .
11. Describe the function and the procedure of the median test.
12. Define the problem of multicollinearity in a simple regression model. How do you detect it and correct it?
13. Write short notes on any two of the following:
 - (a) Systematic sampling.
 - (b) Order statistics.
 - (c) Heteroscedasticity.